

医療ビッグデータの分析と活用のためのシステム環境

(株)パーキンエルマージャパン 田澤 司

現在、医療ビッグデータの蓄積の基盤の整備は着実に進行しており、膨大なデータが再利用可能な形でデータベースに格納され、また日々その量は増えている。データは格納されているだけでは意味はなく、当然これを様々な角度から分析し、医療制度や病院経営の最適化、また臨床研究上の新しい知見の発見などに役立たせ、実際のアクションに結び付けていく必要がある。

多変量解析を含む統計解析手法や機械学習といった高度分析は、R や Python などを始めとするツールにより実行可能である。しかしながら、膨大なデータを前にしたとき、まず行わなければならないのは全体の俯瞰である。全体の傾向を把握してから、細部への深堀を行う必要がある、またデータの品質の確認も必要である。どんな高度な分析手法を適用しても、低品質の入力データでは、その力を発揮できない。

実際は現在のところ、SQL による件数カウント、平均値の算出、その月次や年次などの時間的推移、といった単純な集計に留まるところが多い。さらに集計値を Excel で可視化してレポートを作成する、という作業も一般的だが、決して効率の良いものではない。このため、最近ではデータの可視化に特化した、いわゆる BI ツールが医療ビッグデータの分析でも注目されている。こうしたツールは以下のような機能を要求されている。

- 1.大規模データでも、思考の分断を発生されない高速性。
- 2.IT や分析リテラシーが低い現場スタッフでも簡単に使える操作性
- 3.データベースや各種ファイルなど多様なデータソースへの接続性

こうしたツールの利用により、単純な集計値のみならず、データの分布の確認が行える。実際、医療ビッグデータの分析事例を見ても、分布の確認を行っているケースは意外に少ない。また、医療データにおいては、外れ値も頻繁に観察される。この外れ値がどういう理由で発生したかの確認も重要であるが、面倒な作業である。いずれの問題も、適切なツールを選択することにより大幅な効率化が期待できる。

さらに、カテゴリカルデータを多く含む医療データから新しい知見を得るには、例えば施設ごと、診療科ごと、などの層別化分析が非常に有効である。層別化分析は深い統計知識を必要としないため、一般の病院スタッフでも理解、解釈が容易というメリットもある。Excel でこうした層別軸を変更しながらグラフを作成するのは大変だが、最近のツールはこの辺りの使い勝手が良く、新しい知見の発見に役立つ。

現在、大きな注目を浴びている AI・機械学習に関しては、適用が進展しているデジタルマーケティング等の世界と違って、医療の世界では、結果導出のプロセスが第三者に説明できなければいけない、という大きな問題があるので注意が必要である。医療の世界にブラックボックス的な機械学習は通用しないのである。

ただし、画像診断の世界では Deep Learning などの手法が役立つであろうし、文献や過去の診療記録の検索などにも AI が有効だろう。また、診療の補助、疾病候補のスクリーニング、見逃し等のミスの防止などに適用されていく可能性はあるが、最終判断は人間が行うという点は、当分変わらないと思われる。

また、医療データと一言にいっても、システム発展の歴史的な経緯より、未だに多様なデータが別々のシステムに独立したコード体系で孤立して保存されている場合も多い。こうしたデータの統合にも AI・機械学習が役に立つ可能性は高いと思われる。